

Multi-View Video Packet Scheduling

Laura Toni, *Member, IEEE*, Thomas Maugey, *Member, IEEE*, and Pascal Frossard, *Senior Member, IEEE*

Abstract

In multiview applications, multiple cameras acquire the same scene from different viewpoints and produce correlated video streams. This results in large amounts of highly redundant data. In order to save resources, it is critical to handle properly this correlation during encoding and transmission of the multiview data. In this work, we propose a correlation-aware packet scheduling algorithm for multi-camera networks, where information from all cameras are transmitted over a bottleneck channel to clients that reconstruct multiview images. The scheduling algorithm relies on a new rate-distortion model that captures the importance of each view in the scene reconstruction. We then propose a problem formulation for foresighted optimization of scheduling policies, which adapt to temporal variations in the scene content. Furthermore, we propose a low complexity scheduling algorithm based on a trellis search that selects the subset of candidate packets to be transmitted for optimized reconstruction quality. Simulation results show the gain of our scheduling algorithm when correlation information is used in the scheduler, compared to scheduling policies with no information about the correlation or non-adaptive scheduling policies. We finally show that increasing the optimization horizon in the packet scheduling algorithm improves the transmission performance, especially in dynamic scenarios where the level of correlation varies rapidly with time.

Index Terms

Foresighted packet scheduling, source correlation analysis, multiview streaming, interview correlation, rate-distortion optimization, multimedia communication.

I. INTRODUCTION

Advances in interactive services and 3D television have paved the road to multiview video applications, in which multiple sources acquire and transmit several correlated media streams [1]–[4]. Multimedia wireless sensor networks and multi-camera video systems are typical examples of multiview setups. The flexibility and the interactivity offered by such applications however comes at the price of increased storage/bandwidth requirements. It also requires high computational complexity at encoder and possibly expensive communication between sources. In this context, distributed source coding (DSC), which is based on the Slepian-Wolf (SW) [5] and Wyner-Ziv (WZ) [6] information theory theorems, has gained attention as new coding paradigm, where encoding stays simple and computational complexity is shifted to the decoder. Several efforts have been made to design practical DSC-based communication schemes for correlated information sources [7]–[9]. However, even if DSC permits to reduce bandwidth requirements, high-complexity decoding schemes are generally induced, encoders usually require some a priori information about the correlation between sources, and the application of DSC to many sources rapidly reaches complexity limits. A distributed video coding (DVC) has been proposed in [10] for multiview sequences, where the correlation among adjacent cameras is exploited to construct the side information (SI) more accurately, while a clustered coding strategy has been proposed in [11] for wireless multimedia sensor networks. Even if the coding complexity is reduced in this case, DSC still depends heavily on a good knowledge of correlation at sources. In multiview scenarios, the shape of this correlation has not been properly modeled yet. Thus, although DSC relies on a priori selection on which sources can be used for the generation of SI, the optimization of this selection policy is still an open problem.

In this work, we aim at providing insights on how resource allocation strategies can benefit from correlation awareness in a multi-camera scenario. The proposed framework targets the optimization of scheduling of packets from correlated sources under delay and bandwidth constraints. It can be extended to multi-camera systems with specific source coding techniques (e.g., multiview video coding, DSC, and multiple description coding). However, rather than focusing here on source coding schemes, we are interested in a scenario where each camera independently acquires part of a scene without communication between cameras and without a priori knowledge of correlation between views. The encoded views are then transmitted with a correlation-aware packet scheduling algorithm driven by a wireless access point (AP) (see Fig. 1). The packet scheduling algorithm filters packet to reduce the transmission cost and satisfy the resource constraint in the system. The packets are transmitted then to clients that might independently choose to decode (part of) the 3D scene. In order to optimize the reconstruction quality under bandwidth constraints, one has however to properly select the packets to be transmitted, along with their transmission schedule. Classical rate allocation (RA) techniques cannot solve such a scheduling problem, since they usually do not consider correlation between sources. In this work, we demonstrate the need of optimized correlation-aware scheduling policies in multiview systems, which are able to efficiently share among cameras. Thus, we propose a novel rate distortion (RD) model that estimates the distortion in scene reconstruction from multiple correlated images. Then, we build a technique that minimizes the distortion in the scene reconstruction and adapts the transmission scheme to temporal variations of the scene content.

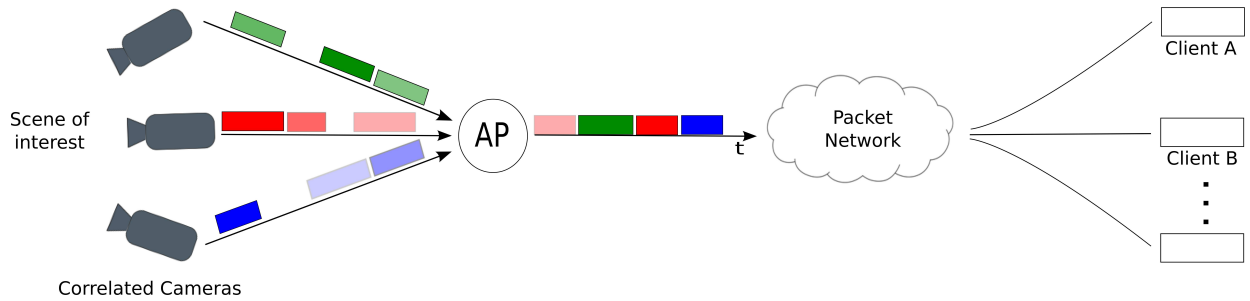


Figure 1. Multi-camera system, with bandwidth bottleneck at the access point.

In particular, we propose a foresighted scheduling algorithm, which is refined at each transmission opportunity. For such an algorithm to reach optimality though, a large time horizon has to be considered in the optimization, which leads to high computational complexity. Thus, we propose a suboptimal trellis-based algorithm able to reduce the complexity while still preserving most of the benefits of foresighted optimization. Simulation results demonstrate that the proposed scheduling algorithm outperforms correlation-agnostic scheduling policies or static camera selection algorithms.

The remainder of this paper is organized as follows. Related works on multiview data gathering are described in Section II. In Section III, some technical preliminaries are given and our new RD model is introduced. The packet scheduling problem is formulated in Section IV and the trellis-based optimization solution is provided in Section V. In Section VI, we discuss the simulation results, and we conclude in Section VII.

II. RELATED WORKS

In this section, we first provide a general overview of the most relevant works from the literature that focus on multi-camera streaming and we highlight the key differences with our work. Then, we describe in more detail the research work in resource allocation and correlation-aware multiview streaming.

In multiview streaming, the prior studies usually address the problem of providing interactivity in selecting views, while saving on transmitted bandwidth and view-switching delay [2], [12]–[16]. The work in [2] is mainly focused on coding views with a minimum level of redundancy in order to simplify the view switching, and the works in [14], [17] optimize the selection of views to be encoded and transmitted based on the user interest. The authors in [15], [18] investigate the transmission of multiview video coded streams on P2P networks and IP multicast, respectively. These works mainly focus on the coding aspects and DSC is often proposed as a solution to reduce encoding complexity [10] or to provide interactive access to the different views [19].

In our work, we consider a general scenario in which the global correlation is not known a priori by each camera and where data gathering targets generic services of providing all views to client sets or proxy servers. Multiple cameras independently capture a 3D scene, but the bandwidth is not sufficient for uploading the full set of captured frames. Therefore, our problem is rather defined as a rate allocation problem. Multi-camera resource allocation solutions in the literature often ignore the dynamic correlation model in multi-camera scenarios and rather focus on optimizing the resources for each camera independently. In other words, they usually optimize the scheduling policy in evaluating the cost, the distortion gain and the time constraints of each camera separately and ignores the possible correlation among cameras. This may result in suboptimal allocation of the network resources. Resource allocation techniques have for example been considered in [20] for video surveillance systems, in which each of the camera captures and transmits the video information in a multihop network. The optimization of the resource allocation (i.e., the time sharing between sources) is based on both the network and source information, but ignores the correlation between the sources.

In a more general resource allocation framework, few works have introduced the sources correlation in the optimization of transmission schemes. In [21], a three-step approach is proposed to optimize the resource allocation between spatially correlated sources for multi-cell frequency-division multiple access (FDMA) networks. However, multimedia transmission is not considered in the optimization. In [22], the level of spatial correlation between sources has been considered at the multiple access control (MAC) layer for wireless sensor networks. In this work, the authors assume that the network needs to estimate an event S . Due to the correlation between neighboring sensors, only part of them might be selected for sending information to the sink, so that the transmission data rate is limited. The MAC protocol prioritizes the access to representative nodes, i.e., nodes with reduced levels of correlation. The same intuition has been considered in [23] and applied to multimedia streaming. A spatial correlation model for visual information in wireless multimedia sensor networks (WMSNs) has been proposed, introducing an entropy-based analytical framework to evaluate the visual information offered by multiple cameras. When the network resources are insufficient the cameras that maximize the joint entropy in a camera set are selected for transmission. The model however only solves a static correlation-based *camera selection* technique, while in our work we consider a dynamic

correlation-based *packet scheduling* optimization. In particular, the framework in [23] can be seen as a particular case of our problem, where both cameras and scene content are static.

The correlation model proposed in [23] has been also used in [24], where the problem of efficient gathering of visually correlated images from multiple sensors has been investigated. The contribution of the author is threefolds: i) the optimal location to place multimedia processing hubs is found in order to ensure the effectiveness of channels frequency reuse; ii) the grouping of cameras into hubs has been optimized in such a way that the joint compression gain is maximized by jointly encoding correlated images associated to the same hub; iii) a scheduling optimization is proposed with the final goal of network lifetime maximization. In particular, each sensor differentially encodes its image conditionally to previously overheard transmissions of broadcaster nodes. The scheduling optimization is aimed at reducing the energy consumption during transmissions by exploiting a correlation-aware differential encoding technique. However, the model is highly sensitive to transmission failures. Moreover, the cameras grouping optimization is based on the assumption of a static correlation, which does not hold in dynamic scenarios. Our work is substantially different from [24], since we propose a packet scheduling optimization that i) is able to adapt to correlation variations in dynamic scenes, ii) considers independent source coding (i.e., it preserves simplicity at the source side), and iii) can be extended to lossy channels.

Finally, it is worth noting that the correlation between cameras might be exploited not only for DSC or resource allocation techniques, but also for error resilience. For example, the correlation between views is implicitly considered in [25], which optimizes interactive multiview streaming over wireless wide area networks (WWAN). A cooperative peer-to-peer repair (CPR) technique is considered to alleviate packet losses in WWANs; the transmission decision of each peer during CPR is optimized based on a Markov decision process (MDP) that exploits the view correlation in data recovery. Even if view recovery through correlated neighbors is considered in [25], the essence of this study is radically different from our work, in which we have no multicast session, no interactive streaming but only multiple cameras that share limited network resources and transmit correlated video streams to a central node.

There are important differences between the above works and the study proposed in this paper. First, we focus our attention on the important problem of optimizing scheduling algorithms such that view correlation is exploited efficiently, without a priori information about the view correlation. Second, even if some other works have investigated resource allocation techniques for multiview scenarios, dynamic view correlation and dynamic packet scheduling solutions are not studied in the literature about multi-camera image capture systems. This is exactly what we propose to address in this paper.

III. FRAMEWORK

We now describe the framework considered in our work. First, we present the multi-camera system and describe the multiview acquisition and transmission process. Then, we introduce the scene reconstruction method and show that the correlation between cameras plays a crucial role in the reconstruction of missing frames at the decoder. Finally, we propose a new rate-distortion model for the representation of the 3D scene information.

A. Multi-camera system

We consider M cameras that acquire images and depth information of a 3D scene from different viewpoints. The images acquired by the M correlated cameras need to be collected by a common AP that eventually transmits (part of) the 3D scene information to clients. Due to bandwidth constraints in the communication system (e.g., on the wireless channel, or on the path between AP and clients), it might not be possible to transmit all the frames from all the cameras to the clients. Thus, at each transmission opportunity, it is important to accurately select which images have to be scheduled and which ones can be sacrificed (i.e., not transmitted), such that the average distortion is minimized. However, depending on the camera arrangement and the scene information, the frames acquired from the different cameras might be correlated in both time and space. First, each camera acquires temporally consecutive frames, which are correlated, especially for static or low-motion 3D scenes: this is the *temporal correlation* in image sequences. Then, neighbouring cameras might acquire overlapping portions of the same scene; this leads to correlated frames due to the *spatial correlation* between multiview cameras. Both the temporal and the spatial correlations might help in reconstructing the overall scene information if some images are missing at the decoder.

We address the frame selection problem as a resource allocation problem that takes into account the level of correlation among cameras in a novel packet scheduling algorithm. We assume a model in which there is no communication among cameras in order to save bandwidth and power. The only minimal information that is known a priori is the position of the cameras, which is possibly updated when cameras change positions in dynamic settings. Along with depth information, each camera is able to estimate its influence on its neighbors and in particular the contribution that it can offer in the reconstruction of neighbor views. We propose below a novel correlation model where each camera can predict the correlation level with neighboring cameras, without global depth information. This local correlation level, which is a set of simple values representing the influence of the camera in the reconstruction of the neighboring ones, is sent by each camera to the scheduling engine.

Then, we consider that each encoded image at a given time instant from a particular camera is packetized into a data unit (DU) and stored in the camera buffer. Each data unit contains texture and depth information about the 3D scene. All the camera DUs are possible candidates for scheduling. We further assume that the transmission on the wireless channel is based on Time

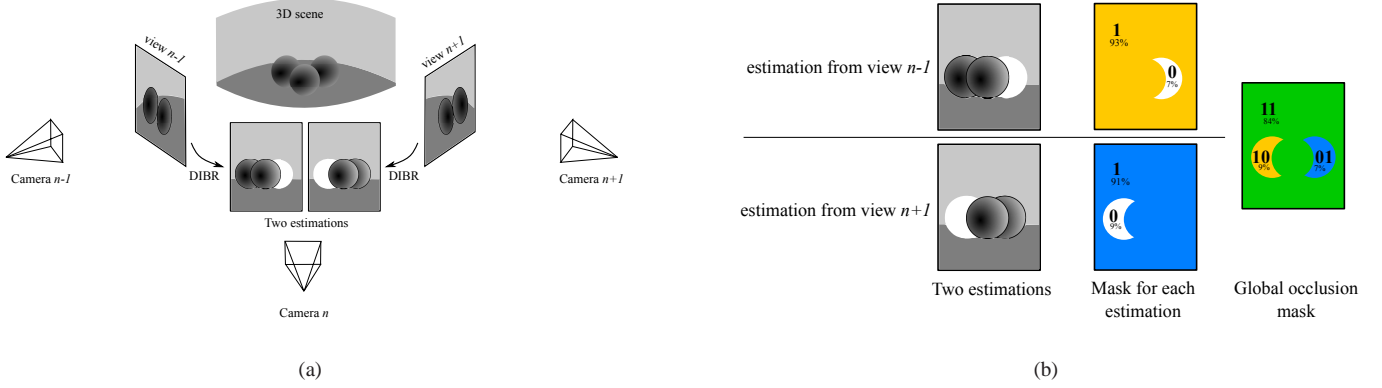


Figure 2. Example of DIBR image estimation at decoder. (a) the central view n is estimated from the two neighboring views $n - 1$ and $n + 1$. (b) the occlusion maps corresponding to the two estimations are merged in order to obtain a global occlusion map with 3 regions. The percentage numbers in the masks indicate the portion of the frame dedicated to each region.

Division Multiple Access (TDMA) model where no more than one DU might be scheduled at any TDMA slot. Once a DU is scheduled for transmission, the channel stays busy for one or multiple time slots, until the current DU has been completely transmitted.¹ Due to streaming delay constraints, the DU needs to be received before a playback deadline, denoted by T_D , in order to be useful for decoding. This means that a DU acquired at the time t stays useful till time $t + T_D$. Data units that have no chance to be received on time are not considered for the scheduling and simply dropped by the cameras. We also assume that the communication channel is lossless such that all the transmitted DUs are correctly received by the access point and subsequently the clients. In this framework, our goal is to propose a correlation-aware scheduling algorithm that selects DUs from different cameras in such a way that the overall distortion in the reconstruction of all camera views is minimized under the bandwidth constraints.

B. Scene Reconstruction

We describe now the scene reconstruction process, which will help to better understand the benefits of exploiting the spatial and temporal correlation of the images. At the receiver side, each frame is decoded independently. The images that have not been transmitted are estimated based on time and/or view interpolation algorithms using information from neighboring frames. More precisely, for the interpolation of a missing view n , the receiver uses images from neighboring cameras with help of depth image based rendering (DIBR) techniques (Fig. 2(a)). Typically, DIBR algorithms use depth information in order to estimate by projection the position of pixels from view k in the missing view n . The projected pixels are generally of good precision (depending on the accuracy of the depth map [26]) but do not cover the whole estimated image, due to visual occlusions. As shown in Fig. 2(b), one can build a binary mask that describes the occluded regions. Then, by merging the estimations obtained by the projections of different neighboring views, we obtain different reconstructed regions in the interpolated image. This can be summarized in a global occlusion map with different regions corresponding to the different occlusions. In the example in Fig. 2(b), the reconstructed scene is subdivided into three regions, each of them is characterized by the set of neighboring views that contribute to the scene reconstruction. In particular, the blue region (which represents 7% of the total scene) is reconstructed based on the estimation from only the view $n + 1$, while for the yellow one (which represents 9% of the total scene) the estimation from view $n - 1$ is considered. The remaining 84% of the scene (i.e., the green region) is reconstructed by merging estimations from both views. The principle for temporal extrapolation is the same. The decoder uses the available past frames to reconstruct a missing frame. The past frames cannot be used to estimate the whole missing image because of occlusions and object motion. The regions where the past frames could give some useful information are computed similarly to the occlusion map in the view interpolation case. The global map with the different prediction regions is used to decide on the best interpolation method for the missing frames at the decoder.

An example of multiview video reconstruction is depicted in Fig. 3, for the case of 8 cameras that acquire several temporally consecutive frames. The goal of the decoder is to reconstruct all the frames in time and space, even if only part of them have been received (red boxes). In the figure, each frame is correlated with frames of the two neighboring views in space, and with the two temporally successive frames (of the same view). If one or more of these correlated frames are missing, the received frames can contribute to the estimation of the missing data (pink boxes). In order to avoid error propagation, only the received frames can be used to reconstruct the missing ones (i.e., reconstructed frames are never used for estimation of other missing frames). Note that we consider temporal estimation only in the forward direction for the sake of simplicity. Our model can

¹From here onwards, we assume the time axis discretized in slots (or scheduling slots) of length equal to the TDMA slot duration.

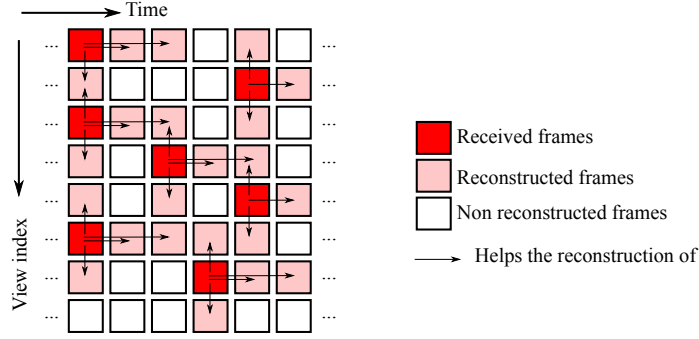


Figure 3. Example of frames reconstruction in multiview video setup. Each frame is correlated with the frames of two neighboring views and with the two successive frames (of the same view). Missing frames are reconstructed from information in the correlated frames that are available at decoder.

however be extended easily to include temporal interpolation in the backward direction too (i.e., from future frames). Finally, a missing frame cannot be reconstructed (white boxes in Fig. 3) when all its correlated frames are missing too.

C. Rate-Distortion Model

We now propose a novel rate-distortion model adapted to the scene reconstruction framework described above. The m -th camera at time t , acquires the image $F_{t,m}$ and compresses it at a rate of $R_{t,m}$ bits per pixel ($m = 1, \dots, M$). A subset of the compressed images captured by all cameras is transmitted to the decoder, which targets the reconstruction of the full scene. If the frame $F_{t,m}$ is available at the decoder, the distortion is directly dependent on the compression or the source rate. If $F_{t,m}$ is missing at decoder, it is reconstructed from the available neighboring frames (in time and space), as described in the previous section.

The overall distortion of the scene at instant t is thus expressed as

$$D_t(\mathbf{R}_t) = \sum_{m=1}^M \frac{1}{w_m} D_{t,m}(\mathbf{R}_t) \quad (1)$$

where w_m represents the relative popularity of a given camera view. The view popularity parameter permits to give a different importance to each camera view in the distortion evaluation (e.g., the central camera might be preferred to the lateral ones). The rate vector \mathbf{R}_t , defined as

$$\mathbf{R}_t = [R_{t,1} \ R_{t,2} \ \dots \ R_{t,M} \ R_{t-1,1} \ \dots \ R_{t-1,M} \ \dots \ R_{t-\rho_t,1} \ \dots \ R_{t-\rho_t,M}]^T,$$

represents the size (in bpp) of the frames received from the different cameras ($m = 1, \dots, M$) in a window of time of size ρ_t , which can be used for the reconstruction of $F_{t,m}$. The distortion $D_{t,m}(\mathbf{R}_t)$ is the distortion of the m -th view at instant t . For each view m acquired at the instant t , we further decompose the frame into regions s_j and we denote by $\mathcal{S}_{t,m}$ the set of such regions. For each $s_j \in \mathcal{S}_{t,m}$, we denote by $\alpha(s_j)$ the relative area of the frame dedicated to the region s_j , such that $\sum_{s_j \in \mathcal{S}_{t,m}} \alpha(s_j) = 1$. In Fig. 2, for example, the frame acquired from the central camera is subdivided in three different regions: the blue, the yellow, and the green ones, with $\alpha(s_j)$ corresponding to 0.07, 0.09, and 0.84 respectively.

Then, a mapping function $\phi_{j,m,t}$ describes which of the neighboring frames can contribute to the reconstruction of the region s_j of the m -th view at time t . In the absence of temporal correlation, the spatially neighboring views only are considered for frame reconstruction. This means that $\phi_{j,m,t} = [\phi_{j,m,t}(1) \ \dots \ \phi_{j,m,t}(M)]$, where $\phi_{j,m,t}(k) = 1$ if the k -th camera is correlated with the region s_j of the frame $F_{t,m}$ and $\phi_{j,m,t}(k) = 0$ otherwise. In this case, \mathbf{R}_t reduces to $\mathbf{R}_t = [R_{t,1} \ R_{t,2} \ \dots \ R_{t,M}]$. When both spatial and temporal correlations are used in the reconstruction, the matrix $\phi_{j,m,t}$ becomes

$$\phi_{j,m,t} = [\phi_{j,m,t}(1) \ \dots \ \phi_{j,m,t}(M) \ \phi_{j,1,t-1}(1) \ \dots \ \phi_{j,m,t-1}(M) \ \dots \ \phi_{j,m,t-\rho_t}(1) \ \dots \ \phi_{j,m,t-\rho_t}(M)]$$

where ρ_t is the number of past frames that can be considered for the reconstruction of the current image. Equipped with the above notation, the distortion $D_{t,m}(\mathbf{R}_t)$ becomes the sum of the distortion in each part s_j^t of the frame at instant t :

$$D_{t,m}(\mathbf{R}_t) = \begin{cases} \sum_{s_j \in \mathcal{S}_{t,m}} \alpha(s_j) d[\phi_{j,m,t} \cdot \mathbf{R}_t] & \text{if the view is not received} \\ d[R_{t,m}] & \text{otherwise.} \end{cases} \quad (2)$$

Finally, the distortion functions $d[R]$ in Eq. (2) can be evaluated from the general expression of the RD function of an intra-coded frame with high-rate assumption [27]:

$$d[R_I] = \mu_I \sigma_I^2 2^{-2R_I} \quad (3)$$

where R_I is the number of bits per pixels and is equal to the sum of the rates that contribute to the current region, σ_I^2 is the spatial variance of the frame and μ_I is a constant depending on the source distribution. It is worth noting that this model has been chosen because it is quite simple and yet accurate. However, our packet scheduling framework is general and other source rate-distortion functions could be used in Eq. (2).

IV. PACKET SCHEDULING ALGORITHM

We discuss in this section a novel packet scheduling framework for wireless multiview camera system that uses the rate-distortion model proposed in the previous section. Then, we propose a novel problem formulation for rate-distortion optimal packet scheduling.

A. Transmission policy

We consider a channel with successive time slots for packet transmission. Each time slot represents a transmission opportunity. The objective is to select which DU should be transmitted at each available timeslot, in order to maximize the quality at the decoder under the playback delay constraint given by T_D . A greedy hence myopic strategy decides the scheduling policy as the one minimizing the overall distortion of the currently acquired frames. However, such a scheduling solution does not necessarily optimize the overall distortion since it does not consider a long term optimization objective. A less myopic scheduling leads the scheduler to allocate more fairly all the views of the camera set with a better global distortion. Thus, in the following we optimize the packet scheduling strategy over a finite time horizon that is generally larger than one transmission time slot.

The delay T_D as well as any temporal parameter introduced in the following is expressed in terms of time slots for the sake of clarity. At the time instant t , all the frames from all the views, acquired in the interval $[t - T_D + 1, t]$ are possible candidates for transmission. They form a set of cardinality L . Let the l -th DU be characterized by its size B_l in bits², its acquisition time slot $T_{A,l}$ (i.e., the instant at which the frame is acquired), its expiration deadline $T_{TS,l} = T_{A,l} + T_D$, and its transmission policy $\pi_l : \{a_l(1) \dots a_l(K)\}$ in the next K timeslots. A transmission policy π_l at time t is a schedule according to which the DU l is allocated for transmission over the time horizon $[t, t + K - 1]$. In particular, $a_l(k) = 1$ means that the data unit l has to be sent at time $t' = t + k - 1$. As the channel is lossless, we assume that each DU is scheduled at most once during its lifetime and that each transmitted DU is sent entirely. Each DU acquired in the time interval $[t - T_D + k + 1, t + k]$ is a possible candidate for transmission at time $t' = t + k, \forall k < K$, when the scheduling policy is optimized at the time instant t . However, the scheduler in general knows only the frames acquired up to the current time t and it is not possible to optimize the scheduling of future frames. Therefore, the scheduling policy has to be refined at each transmission opportunity based on the newly acquired frames. This means that a scheduling policy can change over time. For example, a DU planned for transmission in the future might actually never be transmitted if a future frame with higher importance takes its transmission slot.

We denote by $\pi = [\pi_1 \dots \pi_L]^T$ the scheduling policy for the L candidate DUs. Each policy π leads to a particular distortion on the client side. In this work, we seek the best policy π^* that is able to minimize the expected distortion while satisfying the channel constraints. We formally define below the packet scheduling problem in our new framework.

B. Problem Formulation

We first consider the scheduling problem for a single DU. In this case, the transmission rate is denoted by

$$\mathcal{R}(\pi_l) = B_l \left[\sum_{k=1}^K a_l(k) \right]$$

where $\sum_{k=1}^K a_l(k)$ is equal to 1 if the DU l is scheduled for transmission in the k -th slot, and equal to 0 otherwise. The overall distortion is evaluated as

$$\mathcal{D}(\pi_l, \mathcal{H}) = \begin{cases} D_l(\Psi\{\mathcal{H}\}) & \text{if } \sum_{k=1}^K a_l(k) = 0 \\ D_l(\Psi\{\mathcal{H} \cup l\}) & \text{otherwise} \end{cases} \quad (4)$$

where \mathcal{H} is the set of the DUs already transmitted in the time slots before t (i.e., \mathcal{H} represents the scheduling history), and D_l is the overall distortion level derived from Eq. (2), where the subscripts $\{t, m\}$ have been replaced by the subscript l to describe the data unit l . The function $\Psi\{\mathcal{H}\}$ evaluates the rate vector \mathbf{R} of the M views acquired in the last ρ_t instants given the set of transmitted DUs \mathcal{H} . In particular, each element j of the vector \mathbf{R} is set to B_j if the $j \in \mathcal{H}$, and to 0 otherwise. The evaluation of D_l obviously involves the size and the prediction maps of the data unit, namely B_l and $\{\phi_{j,l}\}$. For the sake of clarity, we omit this dependency in our equations.

²The size of a DU includes the size of both texture and depth data.

We now consider the rate and distortion for multiple DUs. In the joint scheduling of multiple DUs, we evaluate the average distortion and rate for a set of scheduling policies $\boldsymbol{\pi} = [\pi_1 \dots \pi_L]^T$. This outlines the dependency between DUs in the packet scheduling optimization. The average rate for a set of L DUs with a transmission policy $\boldsymbol{\pi}$ is thus given by

$$\mathcal{R}(\boldsymbol{\pi}) = \sum_l \mathcal{R}(\pi_l) = \sum_l B_l \left[\sum_{k=1}^K a_l(k) \right]. \quad (5)$$

The derivation of the average distortion is not as straightforward as the one of the average rate. In particular, the rate of a given DU only depends on the scheduling policy for that DU, while the distortion for a given DU depends on the scheduling policy of the correlated DUs

$$\mathcal{D}(\boldsymbol{\pi}, \mathcal{H}) = \frac{1}{w_l} \sum_{l=1}^L D_l(\Psi\{\mathcal{H} \cup \mathcal{P}_{\boldsymbol{\pi}}\}) \quad (6)$$

where D_l is the distortion for the reconstructed DU l , given the scheduling policy $\boldsymbol{\pi}$, and $\mathcal{P}_{\boldsymbol{\pi}}$ is the set of DUs scheduled in the time slots $[t, t + K - 1]$ based on the scheduling policy $\boldsymbol{\pi}$. Note that, among the DUs in $\mathcal{P}_{\boldsymbol{\pi}}$, the frames correlated with the DU l have an impact in the reconstruction of the l -th DU in the case it cannot be transmitted (i.e., in the case $l \notin \mathcal{P}_{\boldsymbol{\pi}}$).

Equipped with the above definitions of rate and distortion for each policy, we want now to find the best scheduling policy $\boldsymbol{\pi}^*$ that minimizes the average distortion while satisfying the bandwidth constraints. In particular, we seek for

$$\boldsymbol{\pi}^*(\mathcal{H}) = \arg \min_{\boldsymbol{\pi}} \mathcal{D}(\boldsymbol{\pi}, \mathcal{H}) \quad \text{s.t.} \quad \mathcal{R}(\boldsymbol{\pi}) \leq C_{\text{BW}}^* \quad (7)$$

where C_{BW}^* is the bandwidth constraint given by $C \cdot K \cdot T_{\text{TDMA}}$, where C is the channel capacity and T_{TDMA} is the TDMA slot duration in terms of seconds. Due to the dependency among DUs in Eq. (6), the optimization problem can unfortunately not be decomposed easily into mutually independent subproblems. The optimization problem can be solved with exhaustive search methods, which however rapidly become computationally intractable for large time horizon K and large number of cameras M . An alternative solution consists in solving the optimization problem with iterative algorithms, where policies are optimized sequentially. The authors in [28], for example, propose an iterative sensitivity adjustment (ISA) method where, at each iteration, the transmission policy of a single DU is optimized, keeping the other policies fixed. The overall process is then repeated till convergence. Unfortunately, due to multiple dependencies between DUs in our problem, the iterative method does not necessarily reduce the computational complexity compared to an exhaustive search strategy. In the following section, we describe our approximate yet effective solution to determine the best packet scheduling over the time horizon of size K .

V. TRELLIS-BASED OPTIMIZATION SOLUTION

We propose in this section a new trellis-based method for determining the packet scheduling policies. The key idea to limit the computational complexity relies on an effective pruning strategy based on correlation information. We build a trellis in the solution space as follows. We consider the scheduling optimization problem over the time horizon $[t, t + K - 1]$. In the following, we refer to the time instant $(t + k - 1)$ as the k -th scheduling opportunity (or time slot), with $k \in [1, K]$. At the k -th transmission opportunity, the L DUs that are candidates for scheduling are represented by the states (or nodes) $\{S_{k,1}, \dots, S_{k,L}\}$. Then, a direct edge (or branch) from state $S_{k,j}$ to the state $S_{k+1,i}$ represents the decision of scheduling the i -th DU at the $(k + 1)$ -th transmission opportunity, given that the j -th DU has been transmitted during the k -th slot³. A cost B_i is associated to such edge, which corresponds to the size of the i -th DU. For the sake completeness, we also consider, for each time slot, the null state $S_{k,0}$. A branch heading to the null state means that no frame is scheduled, and a zero transmitting rate is associated to this edge. A sequence of branches forms a *path* and all possible paths form a *trellis*. A *full path* is a path connecting a node at the time slot $k = 1$ to a node at the time slot $k = K$. It represents a feasible scheduling policy optimized over a time horizon K as long as the bandwidth constraints are satisfied (i.e., the sum of the sizes of all transmitted DUs is smaller than the channel capacity). The feasible policy with the minimum distortion is the one leading to the best scheduling policy. Note that, since we do not consider packet retransmissions in our system, the transmission state can only appear once on a path for a given packet.

An example of the trellis-based representation is depicted in Fig. 4, where the scheduling policy considers a time horizon of $K = 3$ in a scenario with four cameras. Before starting the frame transmission (i.e., at the time slot 0) no DUs have been acquired and only the null state is available. In the general case, a scheduling policy at the first time slot (i.e., $k = 1$) is represented by a branch going from a specific state $S_{0,i}$ to any possible state $S_{1,j}$, where $S_{0,i}$ is the state associated to the DU previously scheduled at the time slot $t - 1$. The selected scheduling policy is the one that allocates $F_{1,3}$, then $F_{2,1}$, and finally $F_{3,4}$.

As already mentioned above, while the transmitted rate associated to each branch does not depend on the other branches, the average distortion $\mathcal{D}(\boldsymbol{\pi}^*)$ cannot be evaluated separately for each data unit. Because of the correlation between DUs,

³From here onwards, “branch” or “DU” will be used interchangeably, assuming that each branch represents a scheduled DU.

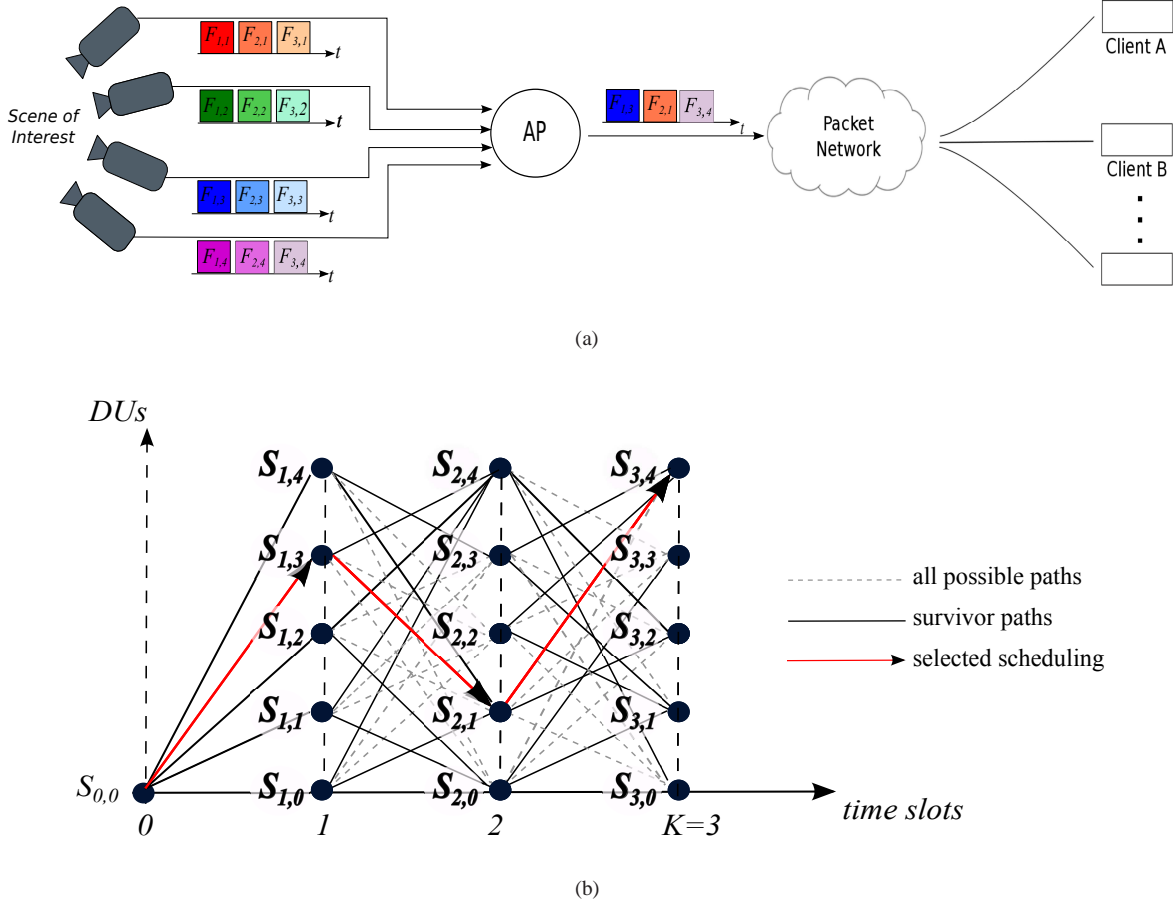


Figure 4. Example of scheduling policy in a scenario of 4 cameras and $K = 3$ transmission timeslots (a) and its associated path in the trellis (b).

the distortion of a given full path is not equal to the summation of the distortion gain of each branch on the path. From an algorithmic point of view, this means that all the branches have to be considered for computing the optimal scheduling solution. Ideally, an exhaustive search should evaluate distortion on all full paths to select the policy with minimum distortion. However, the number of states and full paths are prohibitively large. For example, in a scenario in which L DUs can be scheduled over K time slots, the number of possible full paths is at least $L!/(L-K-1)!$. Rather than an exhaustive search, we propose a suboptimal algorithm that reduces the visited states per time slot and thus substantially reduces the number of full paths to be tested. The key concept is that the best scheduling policy is likely to be the policy that permits the reconstruction of most of the scene. Hence the scheduler shall try to send as much “innovation” as possible, or as little redundancy as possible. Intuitively, once a DU is transmitted, the other DUs that carry correlated information should get a smaller priority. The corresponding branches in the trellis are thus unlikely to be part of the optimal path. Thus, we propose to prune branches depending on the level of correlation that exists between a DU that is candidate for transmission and the set of previously scheduled DUs, denoted by \mathcal{P}_{π_k} where π_k is the scheduling choices (or path) between 1 to k .

In more details, we introduce a branch reward parameter for each branch in the trellis. It is an estimate of the contribution that the DU associated to a given branch can provide to the overall scene reconstruction process, conditioned on the data that have already been scheduled. Consider a given path π_k as the set of DUs scheduled at the first k scheduling opportunities. We evaluate the gain of adding an edge reaching the node $S_{k+1,q}$ to the path π_k : we are interested in the reward of scheduling the DU q at the time slot $k+1$, given that the DUs in the set \mathcal{P}_{π_k} have been previously scheduled. This branch reward is formally given by

$$\rho(S_{k+1,q}|\mathcal{P}_{\pi_k}) = \frac{1}{L} \sum_{l=1}^L \left\{ \sum_{s_j \in F_l} \alpha(s_j) \max \{0, [\phi_{j,l} \cdot \Psi \{ \mathcal{P}_{\pi_k} \cup q \} - \phi_{j,l} \cdot \Psi \{ \mathcal{P}_{\pi_k} \}] \} \right\} \quad (8)$$

In other words, the reward $\rho(S_{k+1,q}|\mathcal{P}_{\pi_k})$ is the “innovative” contribution that the DU q can offer to the reconstructed scene. In particular, for the decoding of the l -th DU among the L DUs under consideration, $\max \{0, [\phi_{j,l} \cdot \Psi \{ \mathcal{P}_{\pi_k} \cup q \} - \phi_{j,l} \cdot \Psi \{ \mathcal{P}_{\pi_k} \}] \}$ is equal to 0 if the region $s_j \in F_l$ can be reconstructed from the previously scheduled DUs (i.e., the DUs in \mathcal{P}_{π_k}), while it is

Algorithm 1 Scheduling Optimization Algorithm

Init: Set $k = 0$. Select all possible branches from the single state in $k = 0$ to all defined states in $k = 1$. Denote by $\{\pi_1\}$ the set of all branches from $k = 0$ to $k = 1$, and by π_1 a generic element of the set.

- 1: **for** $k = 1$ to $K - 1$ **do**
- 2: **for** each path $\pi_k \in \{\pi_k\}$ **do**
- 3: *step a):* for the considered path from 0 to k , individuate all branches going from the scheduling opportunity k to the scheduling opportunity $k + 1$. Denote by \mathcal{B}_{π_k} the set of these branches.
- 4: *step b):* among branches in \mathcal{B}_{π_k} that satisfy the bandwidth constraints identify the subset of the N_s branches with the highest profit $\rho(S_{k+1,q}|\mathcal{P}_{\pi_k})$, with $q \in \mathcal{B}_{\pi_k}$ and discard the remaining branches.
- 5: *step c):* include the N_s selected paths (i.e., the considered path π_k plus the N_s selected branches) in $\{\pi_{k+1}\}$.
- 6: **end for**
- 7: $k \leftarrow k + 1$.
- 8: **end for**
- 9: evaluate the best scheduling policy π_K^* as

$$\pi_K^* = \arg \min_{\pi_K \in \{\pi_K\}} \mathcal{D}(\pi_K) \text{ s.t. } \mathcal{R}(\pi_K) \leq C_{BW}^*$$

equal to 1 if the region cannot be reconstructed from the DUs in \mathcal{P}_{π_k} . In the latter case, the DU q is innovative for the region s_j .

We now describe our solution to optimize the scheduling policy at time t and over a time-horizon of K ; the key concept is that, at each scheduling opportunity, we select of subset of all branches defined in the trellis and we consider the subset as the search space for our packet scheduling policy. The branches in the subset are selected as the ones with the highest branch reward in Eq. (8). We assume that, at time t (i.e., $k = 1$), all branches represent possible candidates for being the first part of the best scheduling solution (i.e., no pruning is done on the first branch of the paths). Thus, we initially determine $\{\pi_1\}$ as the set of branches going from the time slot $k = 0$ (i.e., the node representing the scheduling history) to the time slot $k = 1$. In general, we denote by $\{\pi_k\}$ the set of all paths from $k - 1$ to k (i.e., the set of possible scheduling policies in the first k time slots), and by π_k a generic element of the set. For each path π_k , the search space of possible branches in which the current path can be extended is denoted by \mathcal{B}_{π_k} . From \mathcal{B}_{π_k} , a subset of at most N_s survivor branches are selected as the ones satisfying the bandwidth constraints and maximizing the branch profit $\rho(S_{k+1,q}|\mathcal{P}_{\pi_k})$, with $q \in \mathcal{B}_{\pi_k}$. This means that N_s branches will be considered for constructing the candidate paths π_{k+1} starting from π_k . This subset selection is evaluated for each element in $\{\pi_k\}$, for all the $k > 1$ successively. This leads to at most N_s^{K-1} possible paths for each π_1 . Once the full paths are evaluated, we identify the best scheduling policy as the one that corresponds to the full path minimizing the overall distortion. The overall scheduling algorithm is presented in Algorithm 1. The branch pruning strategy allows us to explore at most $(|\{\pi_1\}|N_s^{K-1})$ paths.

An example of our algorithm is depicted in Fig. 4(b) for a scenario of 4 cameras. In this example, for the sake of simplicity, we assume that the decoding deadline is $T_D = 1$ such that each frame acquired at the time slot k expires at the time slot $k + 1$. We consider the first frame of the sequence and $S_{0,0}$ is the initial state of the scheduler ($t = 1$). No branch is pruned in the first time slot. This means that $\{\pi_1\} = \{(S_{0,0} - S_{1,0}), (S_{0,0} - S_{1,1}), (S_{0,0} - S_{1,2}), (S_{0,0} - S_{1,3}), (S_{0,0} - S_{1,4})\}$, where $(S_q - S_{q'})$ represents the branch going from state S_q to state $S_{q'}$. For each of these branches, we evaluate the full paths as follows. Considering $\pi_1 = (S_{0,0} - S_{1,1})$ and $N_s = 2$, the subset of survivor branches for $k = 2$ is $\{(S_{1,1} - S_{2,4}), (S_{1,1} - S_{2,0})\}$. These two survivor branches are included in $\{\pi_2\}$, and the operation is repeated for every branch $\in \{\pi_1\}$. The branch pruning strategy is considered also for $k = 3$, obtaining then the set $\{\pi_3\}$, which is the set of all the survivor full paths going from $k = 0$ to $k = 3$. In our illustrative example, these paths are represented by solid black lines. Among the candidates full paths, we finally select the best scheduling solution as the one minimizing the distortion as evaluated in Eq. (7).

VI. SIMULATION RESULTS

A. Simulation Setup

We provide now simulation results for a multi-camera scenario where data have to be transmitted over a bottleneck channel of C_{BW} . We consider image sequences where all the DUs from all the cameras have the same size R for the sake of simplicity, and assume that all the views have the same importance, i.e., $w_m = w$ in Eq. (1). Our simulations are carried out with the Ballet and Breakdancer video sequences [29], which consist of $N_f = 100$ frames, at a resolution of $S_R = 768 \times 1024$ pixel/frame and $F_R = 15$ frames per second. The total number of camera views ranges from 4 to 8. We study the performance of our algorithms in different configurations, for different camera setups, different values of the DU size R and for different constraints on the bottleneck bandwidth C_{BW} . In the following, results for both ‘‘Ballet’’ and ‘‘Breakdancer’’ video sequences are provided. It will be noticeable that both sequences lead to very similar results.

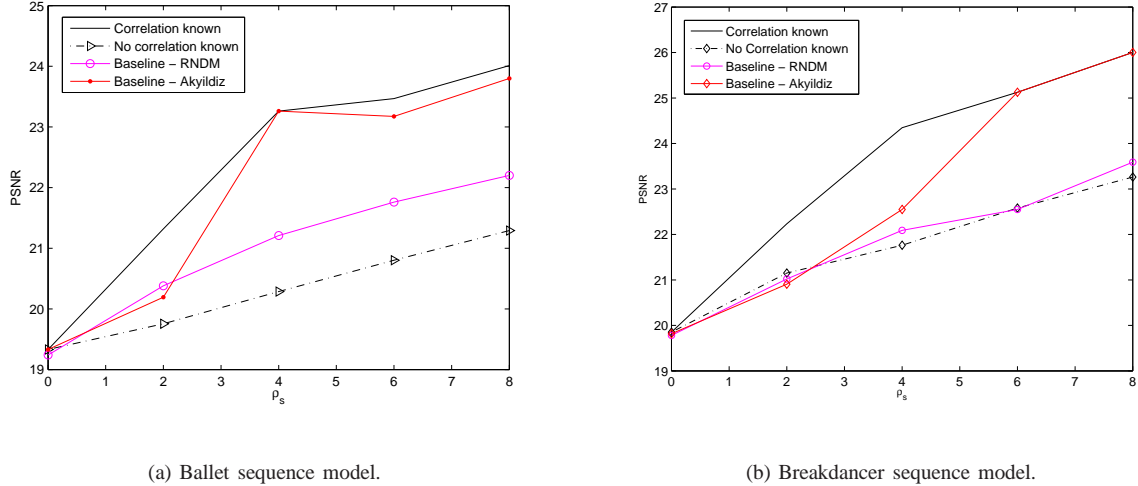


Figure 5. PSNR vs spatial correlation level ρ_s for systems with 8 cameras ($C = 23.5$ Mbps, $r = 11.7$ Mbps, $T_D = 5$, and $\rho_t = 0$).

We denote by ρ_s the number of spatially correlated cameras and we assume that each view is correlated to $\rho_s/2$ neighbor views, if available, on both the left and the right sides. The correlation in time, denoted by ρ_t , is related to the number of frames considered in temporal interpolation at the decoder. We experimentally build the ϕ matrix as explained in Sec. III. To calculate the influence that each camera has on the neighboring ones, we use DIBR techniques and calculate the number of pixels that can be estimated from neighboring views. The resulting matrices ϕ , for both Ballet and Breakdancer video sequences, are available as additional information in the Data Conservatory (DC)⁴. Since we are interested in reconstructing all the views at the decoder, results are provided in terms of mean PSNR, which is the PSNR averaged over all the frames of all views. This means that, even if some frames are decoded at high PSNR values, the average PSNR of the reconstructed scene might be in the low PSNR range in challenging transmission conditions. First, the PSNR of the reconstructed scene is evaluated from the rate-distortion model described in Sec. III-C. Then we validate our findings by experiments with actual reconstruction of the video frames at the decoder. In both cases, we built the temporal correlation model similarly to the spatial one. More precisely, each frame is subdivided into regions, each of them can be reconstructed from frames previously acquired only if no motion occurs in such region. As no motion estimation is employed at the source coding nor at the receiver in our system, only the fixed background contributes to the reconstruction of missing frames.

The proposed algorithm has been compared to two baseline algorithms: a random allocation of the DUs (“Baseline - RNDM”), evaluated averaging the distortion over 1000 runs, and a scheduling solution where cameras priorities are defined a priori based on the joint entropy of the camera dataset as defined in [23] (“Baseline - Akyildiz”). In particular, the camera selection for the latter method is based on the spatial correlation that exists between views, while time correlation information is neglected. The camera priority is established as follows: the camera minimizing the overall distortion becomes the highest priority camera. Then, other cameras are included if they maximize the diversity (i.e., if they minimize the spatial correlation) with respect to the cameras that have been previously selected. We first provide results for a greedy optimization scenario (i.e., $K = 1$) and demonstrate the benefit of a correlation-aware scheduling optimization w.r.t. baseline algorithms. Then we depict the performance of foresighted optimization solutions, showing that low-complexity solutions lead to good performance when the optimization horizon is enlarged.

B. Greedy Optimization

We first analyze the performance of our algorithm in the case where the optimization horizon is limited to the next transmission time slot. We first study the importance of the knowledge of the correlation information in the optimization. Our optimization algorithm is evaluated in different conditions that depend on the type of correlation information considered in the scheduling decisions: i) “Correlation Known”, when the full correlation information is considered in the optimization; ii) “Space Corr Known”, when only the spatial correlation is considered; iii) “Time Corr Known”, when only the temporal correlation is used; iv) “No corr known”, when the scheduler completely ignores the correlation between frames.

We first study the gain that can be achieved when the correlation model is known by the scheduler. In the following figures, the PSNR of the reconstructed scene is evaluated from the rate-distortion model described in Sec. III-C. In the first experiments reported in Fig. 5, the temporal correlation between cameras is neglected both at the scheduler and at the decoder and we focus

⁴This additional material can be downloaded from the DC or ArXiv webpage.

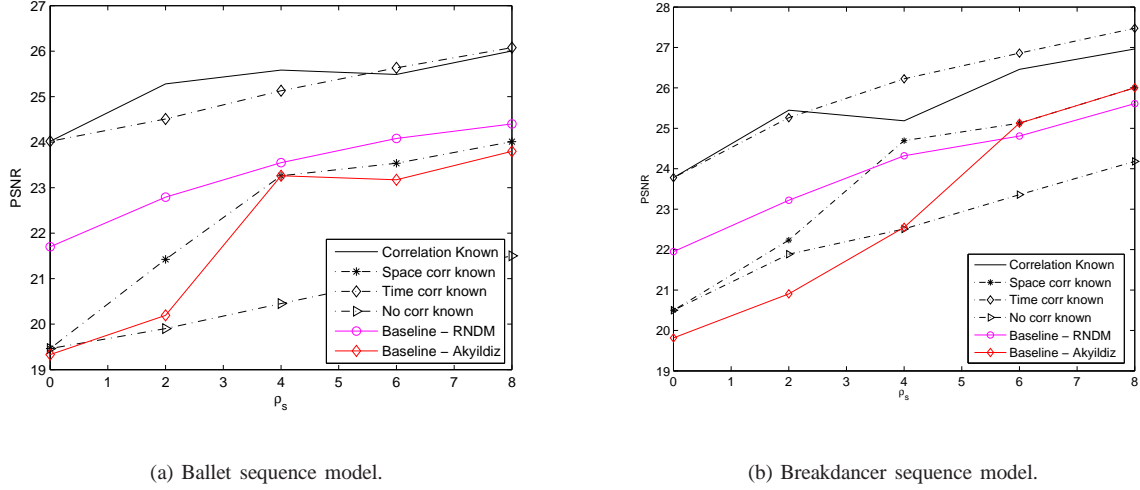


Figure 6. PSNR vs spatial correlation level ρ_s for systems with 8 cameras ($C = 23.5 Mbps$, $r = 11.7 Mbps$, $\rho_t = 3$, and $T_D = 5$).

on the influence of the spatial correlation, which means that missing frames are reconstructed from neighboring views but not from previous frames. The performance of the scheduling algorithm is given as a function of the spatial correlation ρ_s (i.e., a function of the number of views that are considered to be spatially correlated) for systems with 8 cameras, a playback delay $T_D = 5$, a constant encoding rate per camera of $r = 11.7 Mbps$ and a channel capacity $C = 23.5 Mbps^5$. This bandwidth constraint means that 2 only frames out of 8 can be allocated on the channel between each frame acquisition. First, we observe that, for both video sequences, the gain experienced by the algorithm using the spatial correlation information in the scheduling compared to the case in which all the correlation levels are ignored is substantial and this gain increases with the number of correlated frames (i.e., with ρ_s). Thus, the knowledge of the spatial correlation is able to considerably improve the efficiency of the scheduling decisions. Moreover, the proposed algorithm outperforms both baseline algorithms. This means that the packet scheduling optimization leads to a better level of adaptation than the a priori camera selection technique in [23]. It is interesting to note that, by neglecting the correlation model (“No Correlation Known”) the performance becomes very bad and even worse than a random allocation solution, especially for the Ballet video sequence in Fig. 5(a). This means that, rather than choosing the scheduling based on wrong correlation information, it is better to completely ignore it.

In the next experiment, temporal correlation is considered in the scheduling decisions. The PSNR quality is provided in Fig. 6 as a function of the number of spatially correlated cameras ρ_s for systems with 8 cameras, $C = 23.5 Mbps$, $r = 11.7 Mbps$ and a temporal correlation $\rho_t = 3$ (i.e., each frame is considered to be correlated with the three previous frames of the same camera view). It can be observed that the algorithm using temporal correlation (“Time Corr Known”) is the closest one to the algorithm using all the correlation information (“Corr Known”). It has to be noted that all the results provided in the Fig. 6 have been evaluated considering temporal interpolation at the decoder. However, not all the algorithms include this information in the scheduling optimization. For example, the algorithm that only takes into account the spatial correlation information (“Space Corr Known”) is not able to outperform the baseline algorithm with random allocation, especially for the Breakdancer video sequence (Fig. 6(b)), where the random allocation outperforms the “Space Corr Known” strategy for $\rho_s > 2$. This means that, when views are highly correlated in both temporal and spatial domains, a partial information on the correlation does not always lead to a considerable gain in the scheduling optimization. In Table I (Table II), the average PSNR for the sequences reconstructed in the different camera views is provided for the same experiment and for the Ballet (Breakdancer) video sequence. It can be observed that most of the reconstructed camera views achieve the highest PSNR with the correlation-aware scheduling algorithm.

We now repeat similar experiments in a different camera configuration with only 4 views. In Fig. 7, the PSNR quality is measured as a function of the encoding rate ($C = 23.5 Mbps$, $T_D = 5$). It can be observed that there is a tradeoff in the choice of the encoding rate, which varies with the level of correlation information used in the scheduling decisions. This tradeoff is the result of a source quality that increases with encoding rate, while the penalty due to the channel also increases with encoding rate, since more DUs are dropped at high rate for the same channel bandwidth constraint. If there is no known correlation neither in time nor space (i.e., $\rho_s = 0$, $\rho_t = 0$ in Fig. 7(a)), it is better to reduce the encoding rate, so that there is a chance of increasing the number of DUs allocated for transmission, hence the diversity of the information. On the contrary, when the correlation can be exploited both in time and space for frame interpolation (i.e., $\rho_s = 4$, $\rho_t = 2$ in Fig. 7(b)), the

⁵Note that $r = R[bpp] \cdot S_R[\text{pixel per frame}] \cdot F_R[\text{fps}]$.

Table I
AVERAGE PSNR OF THE RECONSTRUCTED IMAGES FOR EACH CAMERA FOR SYSTEMS WITH 8 CAMERAS ($\rho_s = 8$, $\rho_t = 3$, $C = 23.5 Mbps$, $r = 11.7 Mbps$, AND $T_D = 5$), FOR THE BALLET SEQUENCE MODEL.

Optimization Method	Camera view							
	1	2	3	4	5	6	7	8
No Correlation known	24.95	25.32	26.97	27.44	26.88	26.69	25.80	25.26
Correlation known	26.19	26.26	24.13	28.08	26.23	25.18	26.87	26.18
Baseline - Akyildiz	22.28	23.07	24.87	24.52	24.64	25.84	23.84	22.55

Table II
AVERAGE PSNR OF THE RECONSTRUCTED IMAGES FOR EACH CAMERA FOR SYSTEMS WITH 8 CAMERAS ($\rho_s = 8$, $\rho_t = 3$, $C = 23.5 Mbps$, $r = 11.7 Mbps$, AND $T_D = 5$), FOR THE BREAKDANCER SEQUENCE MODEL.

Optimization Method	Camera view							
	1	2	3	4	5	6	7	8
No Correlation known	25.63	25.29	25.58	24.89	24.71	23.80	22.83	22.15
Correlation known	26.69	27.45	27.41	27.47	26.88	28.04	27.23	25.17
Baseline - Akyildiz	25.24	26.34	28.22	24.61	25.13	27.57	26.39	25.70

best encoding rate appears to be a medium rate (17 Mbps). This means that, in this case, rather than scheduling all the frames at low rate (i.e., $r = 5.8 Mbps$), it is better to transmit less frames but at higher rate and to exploit the correlation for the reconstruction of the missing ones. The same qualitative behavior can be observed for the Breakdancer video sequence in Fig. 8, where PSNR quality is measured as a function of the encoding rate.

Finally, we confirm the above observations on experiments with a system that performs actual reconstruction of the video frames at the decoder. These results are provided in Fig. 9. The “Baseline-Akyildiz” performs better than a random scheduling most of the time, but it is in general outperformed by the proposed scheduling optimization, for almost all the values ρ_s of spatial correlation. These observations are in line with our previous results where the quality is measured with the R-D model of Sec. III. They confirm the benefits of including correlation information in the scheduling algorithm, even in a greedy scenario ($K = 1$).

C. Large Optimization Horizon

We now provide results for a framework with foresighted optimization where scheduling policies are computed for several future time slots ($K > 1$). We have already shown above the gain of the proposed algorithm over the baseline ones from $K = 1$, so that we now limit the study to the proposed scheduling algorithm, and look at the gain of a foresighted scheduling policy with respect to a greedy optimization. First, we provide results where the quality is measured with the R-D model of Sec. III (no actual reconstruction of the video frames at the decoder). Then we validate our findings by experiments with actual reconstruction of the video frames at the decoder. For the branch pruning strategy in the trellis-based scheduling solution, we consider the number of survivor branches per time slot to be $N_s = 2$. The results are provided for both a static scenario, where cameras are fixed and the correlation level variations are due to video content, and a dynamic scenario, where cameras are allowed to move in time with a dynamic level of spatial correlation. The random movement of the cameras is simulated as follows. We assume a set of $2M$ possible positions that each camera can take. We start the simulation by randomly allocating each camera in one of the available positions. At each time slot, a camera is randomly selected for changing its position (it

Table III
AVERAGE PSNR OF THE RECONSTRUCTED SEQUENCE FOR EACH CAMERA FOR SYSTEMS WITH 4 CAMERAS ($C = 47 Mbps$, $r = 23.5 Mbps$, AND $T_D = 5$), FOR THE BALLET SEQUENCE MODEL.

Optimization Method	Static Cameras				Moving Cameras			
	$\rho_s = 0, \rho_t = 2$		$\rho_s = 2, \rho_t = 2$		$\rho_s = 0, \rho_t = 2$		$\rho_s = 2, \rho_t = 2$	
	$K = 3$	$K = 5$	$K = 3$	$K = 5$	$K = 3$	$K = 5$	$K = 3$	$K = 5$
Exhaustive search algorithm	24.38	24.55	26.50	26.64	23.13	23.18	25.07	25.20
Branch pruning strategy	24.38	24.52	26.48	26.62	23.12	23.17	25.05	25.18

Table IV
AVERAGE PSNR OF THE RECONSTRUCTED SEQUENCE FOR EACH CAMERA FOR SYSTEMS WITH 4 CAMERAS ($C = 47 Mbps$, $r = 23.5 Mbps$, AND $T_D = 5$), FOR THE BREAKDANCER SEQUENCE MODEL.

Optimization Method	Static Cameras				Moving Cameras			
	$\rho_s = 0, \rho_t = 2$		$\rho_s = 2, \rho_t = 2$		$\rho_s = 0, \rho_t = 2$		$\rho_s = 2, \rho_t = 2$	
	$K = 3$	$K = 5$	$K = 3$	$K = 5$	$K = 3$	$K = 5$	$K = 3$	$K = 5$
Exhaustive search algorithm	23.09	23.25	25.56	25.70	24.18	26.73	24.45	26.92
Branch pruning strategy	23.03	23.23	25.54	25.67	24.18	26.70	24.43	26.91

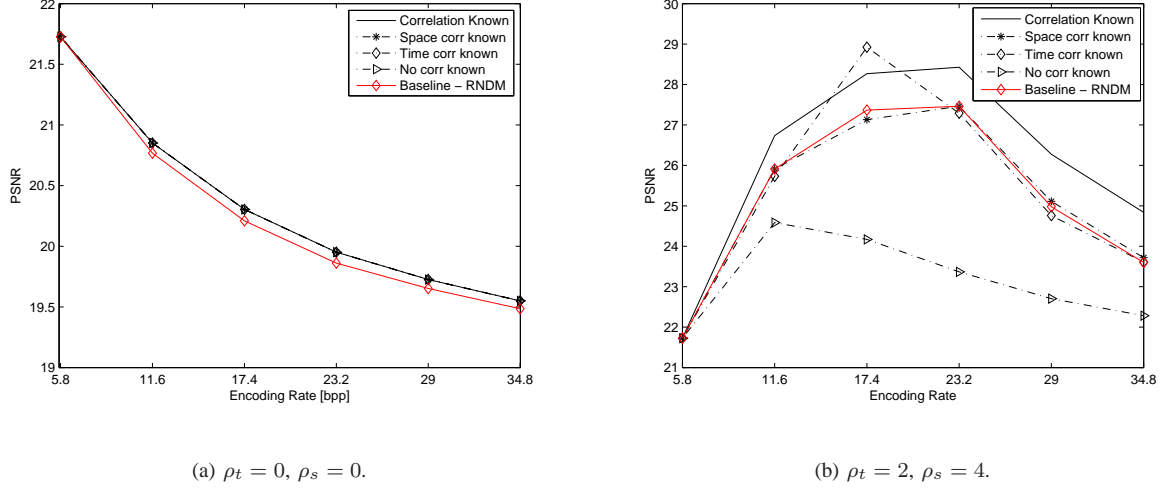


Figure 7. PSNR vs encoding rate for systems with 4 cameras ($C = 23.5 Mbps$, and $T_D = 5$, Ballet sequence model).

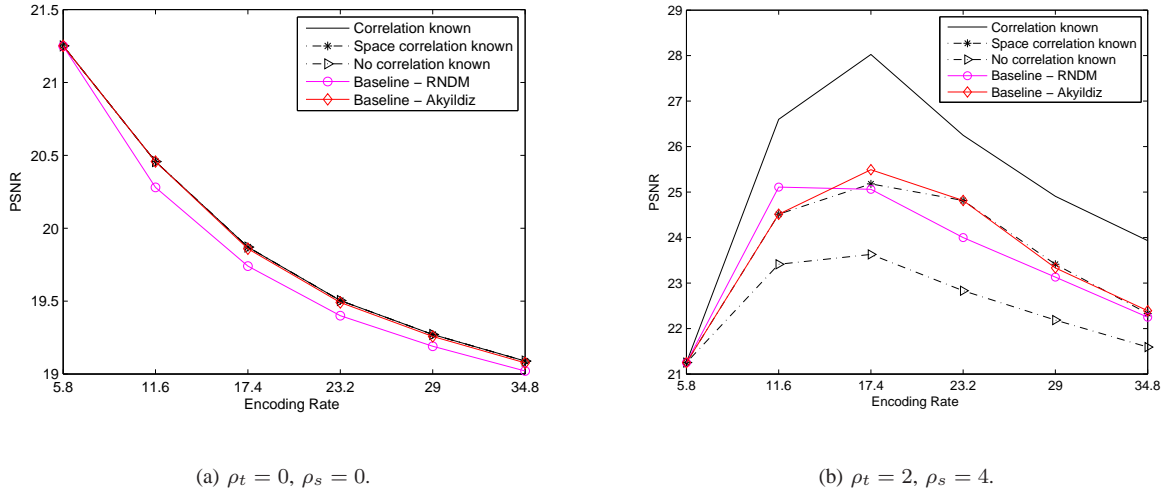
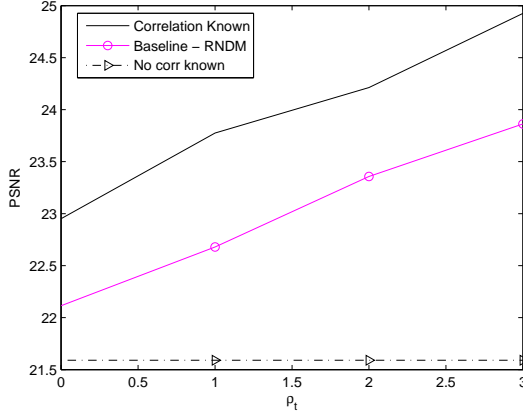


Figure 8. PSNR vs encoding rate for systems with 4 cameras ($C = 23.5 Mbps$, and $T_D = 5$, Breakdancer sequence model).

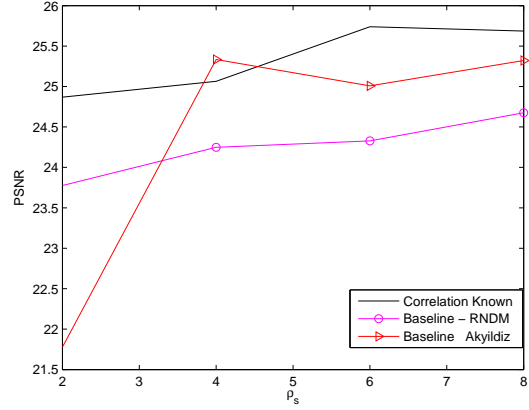
can randomly move to the neighboring position). The camera moves only if the chosen position is not already occupied by another camera; otherwise no movement is performed by the camera set at this time slot. Based on the position of the cameras, the correlation level is evaluated. This means that the correlation between two neighboring cameras can dynamically vary in time, accordingly with the camera movement. In particular, each view can always be reconstructed from the two neighboring ones, but if these two are far apart the portion of frame that can be reconstructed will be small. Moreover, we also assume that the correlation with the frame previously acquired in time is zero when there is a camera motion. Each result provided in the following solution has been averaged over 1000 simulations runs.

We first compare the proposed sub-optimal scheduling algorithm with an optimal one. In particular, we randomly select a time instant $t \in [1, 100]$ and assume that the scheduling history till the time instant $t - 1$ is known⁶. We are interested in optimizing the scheduling policy over a time horizon of K time slots with our trellis-based search technique and with an optimal solution, which exhaustively search for the best scheduling policy. Decoding quality results for the DUs acquired during the time interval under consideration. Results of the reconstructed distortion of the DUs acquired during the time instants $[1, t]$ are provided in Table III and in Table IV, for the Ballet and Breakdancer video sequences, respectively. Each value is averaged over 1000 random simulations for both static and dynamic scenarios with $C = 23.5 Mbps$, $r = 11.7 Mbps$, and $T_D = 5$. It can be observed that, for both sequences, the difference between the branch pruning strategy and the exhaustive search method

⁶The scheduling history is randomly selected.

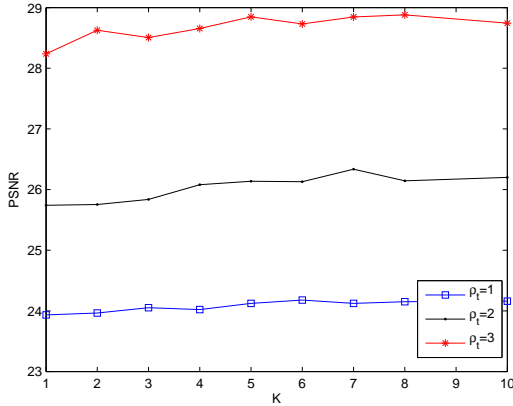


(a) PSNR vs temporal correlation information ρ_t for systems with 4 cameras ($C = 23.5 Mbps$, $r = 23.5 Mbps$, $\rho_s = 2$, and $T_D = 5$).

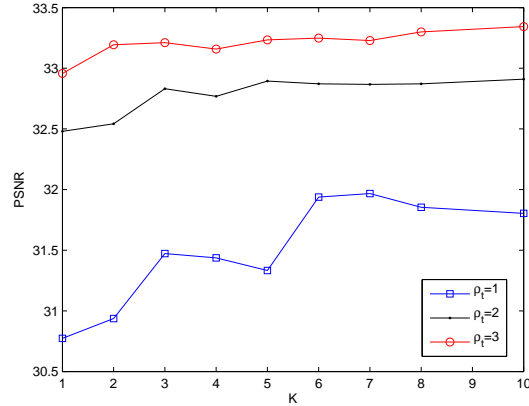


(b) PSNR vs spatial correlation level ρ_s for systems with 8 cameras ($C = 23.5 Mbps$, $r = 11.7 Mbps$, $\rho_t = 3$, and $T_D = 5$).

Figure 9. Reconstructed PSNR for systems with 4 and 8 cameras for different encoding rates and levels of correlation (Ballet sequence).



(a) $C = 23.5 Mbps$



(b) $C = 47 Mbps$

Figure 10. PSNR vs optimization horizon K for systems with 4 dynamic cameras ($r = 23.5 Mbps$, $T_D = 5$, $\rho_s = 2$, and $N_s = 2$, Ballet model sequence).

is negligible. This means that the pruning of the branches in the trellis-based optimization does not penalize significantly the performance, while it drastically reduces the computational complexity.

We now provide results for the proposed foresighted scheduling optimization in dynamic scenarios. In Fig. 10, the model-based reconstructed PSNR is given as a function of the number of optimization time slots K for systems with 4 cameras for several temporal correlation levels ($r = 23.5 Mbps$, $\rho_s = 2$, $C = r$ and $C = 2r$). For all the temporal correlation values ρ_t , we provide results for large K and we observe performance gains with K . Note that the distortion gain due to large K is sometimes marginal for two main reasons: i) the channel capacity is very limited and only few DUs can be scheduled compared to the total number of acquired DUs (Fig. 10(a) where the channel capacity is equal to the source rate of one camera only); ii) there are large levels of correlation so that the system performance is less sensitive to non-optimal scheduling decisions since most of the views will be reconstructed at a fair level anyway (see Fig. 10(b) when $\rho_t = 3$).

In Fig. 11, the PSNR quality is provided as a function of the optimization horizon K for systems with 8 dynamic cameras ($C = 47 Mbps$, $r = 23.5 Mbps$, $T_D = 5$, and $\rho_s = 4$) for both Ballet and Breakdancer video sequences. By increasing the number of cameras from 4 to 8 but keeping the ratio between the channel constraint C and source rate r constant, the number of DUs that cannot be scheduled increases; this makes the selection of the best scheduling policy even more crucial. As expected, the quality gain for large optimization horizons gets more important in this case.

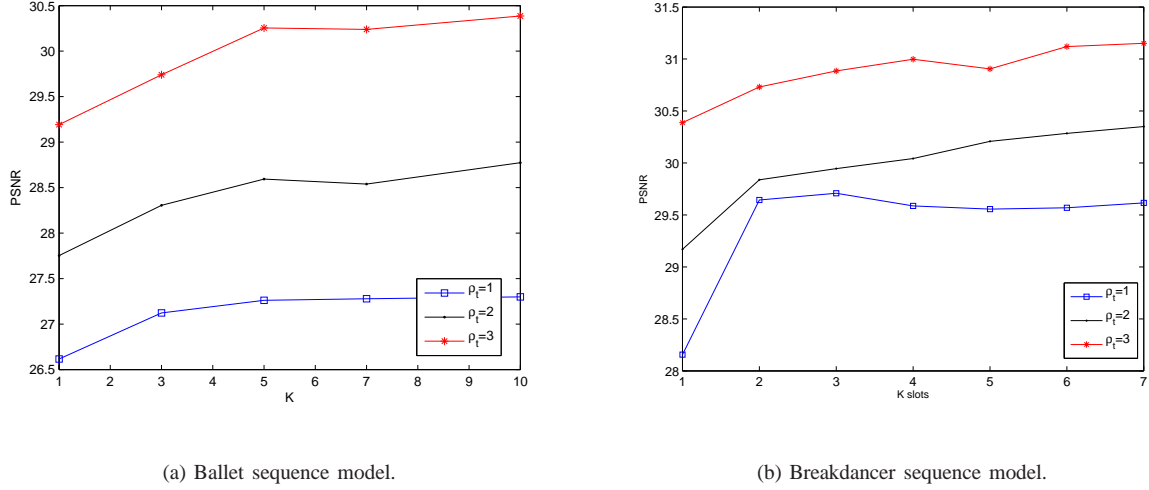


Figure 11. Model-based reconstruction PSNR vs optimization horizon K for systems with 8 dynamic cameras ($C = 47 Mbps$, $r = 23.5 Mbps$, $T_D = 5$, $\rho_s = 4$, and $N_s = 2$).

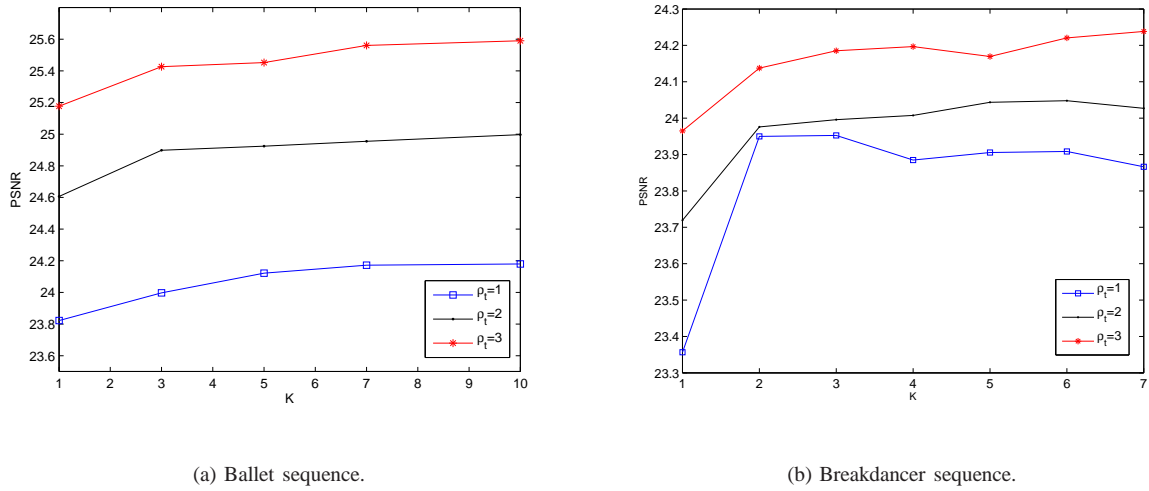


Figure 12. Reconstructed PSNR for systems with 8 dynamic cameras ($C = 47 Mbps$, $r = 23.5 Mbps$, $T_D = 5$, $\rho_s = 4$, and $N_s = 2$).

Finally, in Fig. 12, experimental results are provided for systems with 8 dynamic cameras ($C = 47 Mbps$, $r = 23.5 Mbps$, $T_D = 5$, and $\rho_s = 4$). The experiment is the same of Fig. 11 but the actual reconstruction of the scene is performed at the decoder. As already demonstrated for the greedy optimization results, the *qualitative* behavior of the experimental and model-based results is similar. In general we observe that, the larger the temporal correlation, the better the quality in the reconstruction since more past frames can be used in the reconstruction of a given frame. Furthermore, the experimental results confirm that increasing the optimization horizon improves the performance, as already observed in the results derived from the model-based results.

VII. CONCLUSIONS

We have investigated the impact of frame correlation in the scheduling of packets in a multi-camera system. In particular, we have proposed both a novel RD model able to take into account the correlation level among cameras and a method to estimate the contribution that each camera can offer in the reconstruction of correlated views. Based on this model, we have proposed an optimization algorithm, which determines the packet scheduling policy by taking into account the channel capacity and both the temporal and spatial correlations among encoded views. The proposed algorithm is able to adapt the transmission strategy to the level of correlation experienced by each camera. We have formalized a trellis-based optimization and we have

proposed a suboptimal solution with a tractable complexity, based on effective pruning in a trellis representation. Simulation results have demonstrated the gain of the proposed method compared to classical resource allocation techniques. In addition, it is worth noting that i) when the level of correlation exists in both the time and space domains, knowing at least one of the two correlation levels leads to an improvement in the scheduling algorithm compared to the case where no correlation information is known; ii) the knowledge of the correlation level might help in selecting the best rate at which each camera should encode the images. In particular, the greater the level of correlation, the lower then number of views that needs to be allocated per acquisition time for optimal performance. Finally, we have also demonstrated the robustness of foresighted optimization. As future developments, multiview transmission over lossy channels will be considered in a stochastic optimization framework.

REFERENCES

- [1] W.-P. Yiu, X. Jin, and S.-H. Chan, "VMesh: Distributed segment storage for peer-to-peer interactive video streaming," *IEEE J. Select. Areas Commun.*, vol. 25, no. 9, pp. 1717–1731, Dec. 2007.
- [2] G. Cheung, A. Ortega, and N.-M. Cheung, "Interactive streaming of stored multiview video using redundant frame structures," *IEEE Trans. Image Processing*, vol. 20, no. 3, pp. 744–761, March 2011.
- [3] T. Maugey and P. Frossard, "Interactive multiview video system with non-complex navigation at the decoder," *Accepted on IEEE Trans. Multimedia*, 2012.
- [4] A. Kubota, A. Smolic, M. Magnor, M. Tanimoto, T. Chen, and C. Zhang, "Multiview imaging and 3DTV," *IEEE Signal Processing Mag.*, vol. 24, no. 6, pp. 10–21, Nov. 2007.
- [5] D. Slepian and J. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inform. Theory*, vol. 19, no. 4, pp. 471–480, Jul. 1973.
- [6] A. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inform. Theory*, vol. 22, no. 1, pp. 1–10, Jan. 1976.
- [7] Z. Xiong, A. Liveris, and S. Cheng, "Distributed source coding for sensor networks," *IEEE Sig. Proc. Mag.*, vol. 21, no. 5, pp. 80–94, Sept. 2004.
- [8] Y. Wu, V. Stankovic, Z. Xiong, and S.-Y. Kung, "On practical design for joint distributed source and network coding," *IEEE Trans. Inform. Theory*, vol. 55, no. 4, pp. 1709–1720, Apr. 2009.
- [9] S. Li and A. Ramamoorthy, "Networked distributed source coding," in *Theoretical Aspects of Distributed Computing in Sensor Networks*. Springer Berlin Heidelberg, 2011, pp. 191–224.
- [10] N. Anantrasirchai, D. Agrafiotis, and D. Bull, "Distributed video coding for wireless multi-camera networks," in *Proc. IEEE Int. Conf. on Visual Information Engineering*, Aug. 2008, pp. 67–70.
- [11] P. Wang, R. Dai, and I. Akyildiz, "A spatial correlation-based image compression framework for wireless multimedia sensor networks," *IEEE Trans. Multimedia*, vol. 13, no. 2, pp. 388–401, Apr. 2011.
- [12] E. Kurutepe, M. Civanlar, and A. Tekalp, "Client-driven selective streaming of multiview video for interactive 3DTV," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 11, pp. 1558–1565, Nov. 2007.
- [13] G. Cheung, V. Velisavljevic, and A. Ortega, "On dependent bit allocation for multiview image coding with depth-image-based rendering," *IEEE Trans. Image Processing*, vol. 20, no. 11, pp. 3179–3194, Nov. 2011.
- [14] Y. Liu, Q. Huang, S. Ma, D. Zhao, and W. Gao, "RD-optimized interactive streaming of multiview video with multiple encodings," *Journal of Visual Commun. and Image Representation*, vol. 21, no. 5, pp. 523–532, March 2010.
- [15] E. Kurutepe and T. Sikora, "Multi-view video streaming over P2P networks with low start-up delay," in *Proc. IEEE Int. Conf. on Image Processing*, Oct. 2008, pp. 3088–3091.
- [16] Z. Chen, M. Zhang, L. Sun, and S. Yang, "Delay-guaranteed interactive multiview video streaming," in *Proc. IEEE Int. Symposium on Circuits and Systems*, May 2009, pp. 1795–1798.
- [17] Z. Pan, Y. Ikuta, M. Bandai, and T. Watanabe, "User dependent scheme for multi-view video transmission," in *Proc. IEEE Int. Conf. on Communications*, June 2011, pp. 1–5.
- [18] J.-G. Lou, H. Cai, and J. Li, "Interactive multiview video delivery based on IP multicast," *Journal of Adv. MultiMedia*, vol. 2007, no. 1, Jan. 2007.
- [19] N.-M. Cheung, A. Ortega, and G. Cheung, "Distributed source coding techniques for interactive multiview video streaming," in *Proc. Picture Coding Symp.*, May 2009, pp. 1–4.
- [20] H.-P. Shiang and M. van der Schaar, "Information-constrained resource allocation in multicamera wireless surveillance networks," *IEEE Trans. on Circuits and Syst. for Video Tech.*, vol. 20, no. 4, pp. 505–517, April 2010.
- [21] D. Bandari, G. J. Pottie, and P. Frossard, "Correlation-aware resource allocation in multi-cell networks," *arXiv:1112.2723*, vol. abs/1112.2723, 2011.
- [22] M. C. Vuran and I. F. Akyildiz, "Spatial correlation-based collaborative medium access control in wireless sensor networks," *IEEE/ACM Trans. Netw.*, vol. 14, pp. 316–329, Apr. 2006.
- [23] R. Dai and I. Akyildiz, "A spatial correlation model for visual information in wireless multimedia sensor networks," *IEEE Trans. Multimedia*, vol. 11, no. 6, pp. 1148–1159, Oct. 2009.
- [24] P. Wang, R. Dai, and I. F. Akyildiz, "Visual correlation-based image gathering for wireless multimedia sensor networks," in *Proc. IEEE Int. Conf. on Computer Communications*, Apr. 2011, pp. 746–749.
- [25] Z. Liu, G. Cheung, and Y. Ji, "Distributed markov decision process in cooperative peer recovery for WWAN multiview video multicast," Nov. 2011, pp. 1–4.
- [26] K. Müller, P. Merkle, and T. Wiegand, "3D video representation using depth maps," *Proc. IEEE*, vol. 99, pp. 643–656, April 2011.
- [27] T. A. Cover and J. A. Thomas, *Elements of Information Theory*, 1st ed. New York, NY, 10158: John Wiley & Sons, Inc., 1991.
- [28] P. Chou and Z. Miao, "Rate-distortion optimized streaming of packetized media," *IEEE Trans. Multimedia*, vol. 8, no. 2, pp. 390–404, April 2006.
- [29] [Online]. Available: <http://research.microsoft.com/en-us/um/people/sbkang/3dvideodownload/>